On the Density of Happy Numbers

J. Gilmer

Department of Mathematics, Rutgers, The State University of New Jersey, 110 Frelinghuysen Road Piscataway, NJ 08854

Abstract

The happy function $H : \mathbb{N} \to \mathbb{N}$ sends a positive integer to the sum of the squares of its digits. A number x is said to be happy if the sequence $\{H^n(x)\}_{n=1}^{\infty}$ eventually reaches one. A basic open question regarding happy numbers is what bounds on the density can be proved. This paper uses probabilistic methods to reduce this problem to experimentally finding suitably large intervals containing a high (or low) density of happy numbers as a subset. Specifically we show that $\bar{d} > .18577$ and $\underline{d} < .1138$. We also prove that the asymptotic density does not exist for several generalizations of happy numbers.

Keywords: happy numbers, density, local limit theorem

1. Introduction

It is well known that if you iterate the process of sending a positive integer to the sums of the squares of its digits, you eventually either arrive at 1 or the cycle

$$4 \to 16 \to 37 \to 58 \to 89 \to 145 \to 42 \to 20 \to 4$$

If we change the map, instead sending an integer to the sum of the cubes of its digits, then there are 9 different cycles that are possible (see [5]). There are many possible generalizations of these kinds of maps, for instance Grundman and Teeple in [2] considered the map which sends an integer n to the sum of the e'th power of its base b digits. When appropriate we adopt their notation of (e, b)-happy numbers, which are numbers eventually arriving at 1 under this map. The results of this paper, however, apply to a broader class of functions.

Definition 1.1. *Let $b > 1$ be an integer, and let h be a sequence of $b - 1$ non-negative integers s.t. $h(0) = 0, h(1) = 1$. We define the generalized b-happy function with digit sequence h as the extension $H : \mathbb{Z}^+ \to \mathbb{Z}^+$ where for $n \in \mathbb{Z}^+$ with base b representation $n = \sum_{i=0}^{k} a_i b^i$, $H(n) \equiv \sum_{i=0}^{k} h(a_i)$.*

Email address: jmgilmer@math.rutgers.edu (J. Gilmer)

Fix a generalized b-happy function H. Let $\alpha = \max_{i=0,\ldots,b-1} \big(H(i)\big)$. If n is a d digit integer (in base b), then

$$H(n) \leq \alpha d$$

If d^* is the smallest $d \in \mathbb{N}$ s.t. $\alpha d < b^{d-1}$, then for all n with at least d^* digits, $H(n) < n$. This implies the following

Fact 1.2. *$\forall n \in \mathbb{N}, \exists i$ s.t. $H^i(n) \in [0, b^{d^*-1} - 1]$.*

Moreover finding all the possible cycles for a generalized b-happy function H amounts to performing a computer search on the trajectories of the integers in the interval $[0, b^{d^*-1} - 1]$.

Guy asks in [6, problem E34] a number of questions regarding happy numbers and their generalizations, including the existence (or not) of arbitrarily long sequences of consecutive happy numbers and what bounds on the density of integers arriving at various cycles can be proved. To date, there have been a number of papers in the literature addressing the former question, for example [1],[2],[3]. In regards to the latter question the main result of this paper will allow one to prove (using computer calculations) a lower (upper) bound on the upper (lower) densities for any choice of H as in definition (1.1), and any cycle.

Our results only give one sided bounds. In an earlier version of this manuscript (posted on the arxiv), we asked, in the case of $(2, 10)$-happy numbers, if $\bar{d} < 1$. Recently Moews [6] has announced a proof of this, and along the way gives an alternate proof of the fact that the density does not exist. Specifically, he shows that $.1962 < \bar{d} < .38$, and $0.002937 < \underline{d} < .1217$.

2. Basic Definitions

Remark 2.1. *Unless otherwise noted, we regard an interval $I = [a, b]$ as the integer interval $\{n \in \mathbb{Z}^+ \mid a \leq n \leq b\})$ where in general $a, b \in \mathbb{R}$. As usual we denote $|I|$ to be the cardinality of this set. We denote $[n]$ as the set $\{0, 1, \ldots, n\}$.*

Definition 2.2. *Let I be an integer interval and Y the random variable uniformly distributed amongst the set of integers $\subset I$. Then we say the random variable Y is induced by the interval I.*

Definition 2.3. *Let H be a generalized b-happy function. Let $C \subseteq \mathbb{N}$. If for an integer n, $\exists k$ s.t. $H^k(n) \in C$, then we say n is type-C.*

Definition 2.4. *We say an integer interval I has type-C density d if*

$$\frac{|\{n \in I \text{ s.t. } n \text{ is type-}C\}|}{|I|} = d$$

Remark 2.5. *If Y is the random variable induced by the interval I then the type-C density of I equals $\mathbb{P}\big(H(Y) \text{ is type-}C\big)$.*

Usually the set C will be assumed to be a cycle we are interested in. However in the case we wish to prove bounds on the lower densities, we instead find the upper density of the set of type-C' integers, where C' is taken as the union of all the possible cycles except for C.

2.1. The random variable $H(Y_m)$

Consider the random variable Y_m induced by the interval $[0, b^m - 1]$, i.e. Y_m is a random m-digit number. If X_i is the random variable corresponding to the coefficient of b^i in the base-b expansion of Y_m, then

$$H(Y_m) = \sum_{i=0}^{m-1} H(X_i) \tag{1}$$

In this paper we will be interested in the mean and variance of $H(Y_1)$ (i.e. the image of a random digit) which we refer to as the *digit mean* (μ) and *digit variance* (σ^2) of H. The $H(X_i)$ in (1) are all i.i.d random variables, thus

$$\mathbf{E}[H(Y_m)] = \mu m, \ \ \mathbf{Var}[H(Y_m)] = \sigma^2 m \tag{2}$$

(1) implies a number a facts about $H(Y_m)$. $H(Y_m)$ is equivalent to rolling m times a b-sided die with faces $= 0, 1, H(2), \ldots, H(b-1)$ and taking the sum. It approaches a normal distribution as m gets large. Also the distribution of $H(Y_m)$ is concentrated near the mean which implies the following key insight:

Remark 2.6. *The density of happy numbers amongst m digit integers depends almost entirely on the distribution of happy numbers near μm.*

2.2. Computing Densities

Let $P_{m,i} = \mathbb{P}\big(H(Y_m) = i\big)$. Then

$$P_{m,i} = \frac{\left|\{(a_1, a_2, \ldots, a_m) \mid a_k \in H([b-1]) \text{ and } \sum\limits_{k=1}^{m} a_k = i\}\right|}{b^m}$$

For fixed m, the sequence $(P_{m,i})$ has generating function

$$f_m(x) = \sum_{i=0}^{\infty} P_{m,i} x^i = \left(\frac{1 + x + x^{H(2)} + \cdots + x^{H(b-1)}}{b}\right)^m \tag{3}$$

This implies the following recurrence relation with initial conditions $P_{0,0} = 1$, $P_{0,i} = 0, i \in \mathbb{Z} - \{0\}$.

$$P_{m,i} = \frac{P_{m-1,i} + P_{m-1,i-1} + P_{m-1,i-H(2)} + \cdots + P_{m-1,i-H(b-1)}}{b} \tag{4}$$

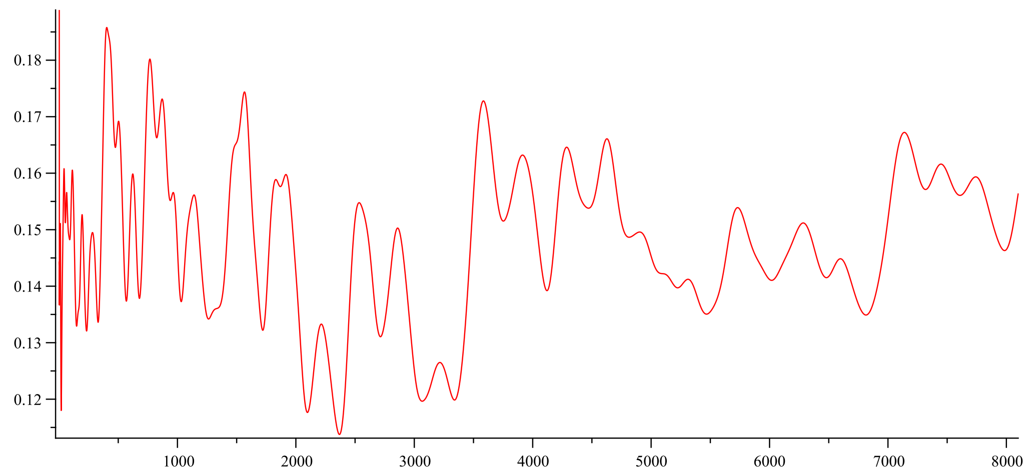
To see this, write $f_m(x) = \left(\frac{1+x+x^{H(2)}+\cdots+x^{H(b-1)}}{b}\right)^{m-1} \left(\frac{1+x+x^{H(2)}+\cdots+x^{H(b-1)}}{b}\right)$ and consider the coefficient of x^i.

If $\alpha = \max\limits_{i=0,\ldots,b-1} \big(H(i)\big)$, then $H(Y_m) \subseteq [0, m\alpha]$, and in particular $P_{m,i} = 0$ if $i > m\alpha$. Using this fact combined with (4) we can implement the following simple algorithm for quickly calculating the type-C density of the integer interval $[b^m - 1]$.

1. First using the recurrence (4), calculate $P_{m,i}$ for $i = 0, \ldots, m\alpha$.
2. Using brute force, find the type-C integers in the interval $[0, m\alpha]$.
3. Output $\sum_{\substack{i \in [0,m\alpha] \\ i \text{ type-C}}} P_{m,i}$.

Using this algorithm, calculating the density for large n becomes computationally feasible. The Figure 1 graphs the density of happy numbers $< 10^n$ for n up to 8000.

Figure 1: Relative Density of Happy Numbers $< 10^n$



The peak near 10^{400} and valley near 10^{2350} will be used to imply the bounds obtained in this paper.

2.3. A Local Limit Law

The random variable $H(Y_m)$ approaches a normal distribution as m becomes large. The following theorem[1] presented in [4, p. 593] gives a bound.

Theorem 2.7. *(Local limit law for sums). Let $X_1, \ldots, X_n$ be independent integer valued variables with PGF $B(z)$, mean μ and variance σ^2 where it is assumed that the X_i are supported on $\mathbb{Z}^+$. Assume that $B(z)$ is analytic in some disc that contains the unit disc in its interior and that $B(z)$ is aperiodic with $B(0) \neq 0$. Then the sum,*

$$S_n := X_1 + X_2 + \cdots + X_n$$

satisfies a local limit law of the Gaussian type: for t in any finite interval, one has

$$\mathbb{P}(S_n = \lfloor \mu n + t\sigma\sqrt{n} \rfloor) = \frac{e^{\frac{-t^2}{2}}}{\sqrt{2\pi n}\sigma} \left(1 + O(n^{\frac{-1}{2}})\right)$$

[1] We quote a simpler version, with a minor typo corrected

.

Here aperiodic means that the $\gcd\{j \mid b_j > 0, j > 0\} = 1$, where $B(z) = \sum_{j=0}^{\infty} b_j z^j$. In our case the probability generating function (PGF) of the $H(X_i)$ is the polynomial

$$p(x) = \frac{x^{H(0)} + x^{H(1)} + \cdots + x^{H(b-1)}}{b}$$

Thus since we assume $H(0) = 0, H(1) = 1$ the above theorem applies for the sum $H(Y_m)$. As a consequence, for a fixed interval $[-T, T]$, if $i = \mu m + t\sigma\sqrt{m}$ for some $t \in [-T, T]$ then

$$P_{m,i} = \frac{e^{\frac{-t^2}{2}}}{\sqrt{2\pi m}\sigma}\left(1 + O\left(m^{\frac{-1}{2}}\right)\right)$$

The constant for the error term $O(m^{\frac{-1}{2}})$ is dependent on the choice of T, for our purposes we do not need an explicit constant.

2.4. A Heuristic

Remark 2.8. *(General motivation for the proofs) Suppose I is a large interval containing type-C integers with some density d. Consider the choices of m s.t. the mean of $H(Y_m)$ is in the interval I, then for some choices of m we likely have*

$$\mathbb{P}\big(H(Y_m) \text{ is happy}\big) \geq d$$

This of course is just a heuristic, and we will only be able rigorously prove the weaker result, $\exists m$ s.t. $\mathbb{P}\big(H(Y_m) \text{ is happy}\big) \geq d\big(1 - o(1)\big)$. The key idea for the proof is to average over all reasonable choices of m in order to imply there is an m with the desired property.

Our main tool will be using theorem (2.7) to say that for small k, $H(Y_m)$ and $H(Y_{m+k})$ have essentially the same distribution only shifted by a factor of μk. Thus as m varies, the distributions of the $H(Y_m)$ should uniformly cover (with small error) the interval I. This argument will not work immediately, as the fuzzy term in the local limit law prevents us from obtaining explicit bounds on the error (and any explicit bounds seem unsatisfactory for our purposes). Section 3 adds a necessary step, which is to construct an interval within $[b^{n-1}, b^n]$ with high type-C density for n arbitrarily large. The main result is presented in section 4, the proof uses the local limit law with the result from section 3.

3. Constructing Intervals

Definition 3.1. *We say an integer interval I is n-strict if $I \subseteq [b^{n-1}, b^n - 1]$, and $|I| = b^{\frac{3n}{4}}$.*

The primary goal of this section is to construct n-strict intervals of high type-C density for arbitrarily large n.

Our choice of the definition of n-strict is only for the purpose of simplifying calculations, there is nothing special about the value $\frac{3}{4}$, in fact any ratio $> \frac{1}{2}$ would work. Note that if 4 does not divide n, then no n-strict intervals exist.

For the entirety of this section we will make the following assumptions:

- H is a generalized b-happy function with digit mean μ and digit variance σ^2.
- We are interested in the upper density of type-C integers for some $C \subset \mathbb{N}$.
- We have found via computer search an appropriate starting interval I_1 which for some $n_1 \in \mathbb{N}$, is n_1-strict and has type-C density $= d_1$.

The results in this section only apply if this n_1 is sufficiently large, so we state here exactly how large n_1 must be so one knows where to look for the interval I_1. In particular we say an integer n satisfies bound (**B**) if

B1: $4\left(1 + 3\mu + \sqrt{2}\sigma b^{\frac{5n}{8}}\right) \leq b^{n-1}$

B2: $\sqrt{3\mu b}\sigma \leq b^{\frac{3n}{8}}$

B3: $4\mu\left(3\mu + 1 + b^{\frac{3n}{4}} + 2\sigma\mu^{\frac{-1}{2}} b^{\frac{5n}{8}}\right) \leq b^{n-1}$

For most of the interesting cases of b-happy functions, for example the cases (e, b) s.t. $e \leq 5$, assuming $n > 23$ is enough to guarantee that it satisfies bound (**B**). This is well within the scope of the average computer, as it is possible to calculate the densities for n up to (and beyond) 1000 using the algorithm in section 2. These bounds are necessary in the proof of theorem (3.5).

Our first goal is to use an arbitrary n-strict interval I to construct a second interval I_2 which is n_2-strict for some n_2 much larger than n and contains a similar density of type-C integers as I. The next lemma will be a helpful tool.

Lemma 3.2. *Let $I = [i_1, i_2]$, $J = [j_1, j_2]$ be integer intervals. Let $S \subseteq I$ and Y a random variable taking integer values in J. For $k \in \mathbb{Z}$ denote $\tau_k(Y)$ as the random variable $Y + k$. Then there exists a (possibly negative) integer $k \in [i_1 - j_2, i_2 - j_1]$ s.t. $\mathbb{P}\big(\tau_k(Y) \in S\big) \geq \frac{|S|}{|I|+|J|-1}$.*

Proof. The idea of the proof is that by averaging over all appropriate k, the distributions of $\tau_k(Y)$ should uniformly cover I. More formally, let $k_1 = i_1 - j_2$, $k_2 = i_2 - j_2$, and let K be the set of integers in the interval $[i_1 - j_2, i_2 - j_1]$, which we denote as $\{k_1, k_1 + 1, \ldots, k_2\}$. Note that $|K| = |I| + |J| - 1$. Pick k uniformly at random from K, and consider the random variable $Z = Pr\big(\tau_k(Y) \in S\big)$. Then

$$\mathbf{E}[Z] = \frac{1}{|K|} \sum_{k=k_1}^{k_2} \mathbb{P}\big(\tau_k(Y) \in S\big)$$

$$= \frac{1}{|I|+|J|-1} \sum_{k=k_1}^{k_2} \sum_{i \in S} \mathbb{P}\big(\tau_k(Y) = i\big)$$

$$= \frac{1}{|I|+|J|-1} \sum_{i \in S} \sum_{k=k_1}^{k_2} \mathbb{P}(\tau_k(Y) = i) \tag{5}$$

Note that $\mathbb{P}\big(\tau_k(Y) = i\big) = \mathbb{P}\big(Y = i - k\big)$. And for $i \in I$, $i - k_1 \leq j_2$, $i - k_2 \geq j_1$. Thus $\forall i \in S$, $\sum\limits_{k=k_1}^{k_2} \mathbb{P}\big(\tau_k(Y) = i\big) = 1$. Therefore

$$(5) = \frac{|S|}{|I|+|J|-1}$$

So $\exists k$ s.t $Pr\big(\tau_k(Y) \in S\big) \geq \mathbf{E}[Z] = \frac{|S|}{|I|+|J|-1}$ proving the claim.

□

Using lemma (3.2) we will not lose much density assuming that $|I|$ is much larger than $|J|$. However when considering the random variable $H(Y_m)$ the set $|J|$ will be much too large. As a result it will be more useful to consider a smaller interval where the bulk of the distribution lies.

Lemma 3.3. *Let Y be an integer valued random variable with mean μ_Y and variance σ_Y^2, and let $\lambda > 0$. Let S be a set of integers $\subseteq [i_1, i_2] = I$ with density d. Then $\exists k$ an integer $\in [i_1 - (\mu_Y + \sigma_Y \lambda), i_2 - (\mu_Y - \sigma_Y \lambda)]$ s.t. $Pr\big(\tau_k(Y) \in S\big) \geq (1 - \frac{1}{\lambda^2})(\frac{d}{1+\frac{2\sigma_Y\lambda}{|I|}})$.*

Proof. By Chebyshev's Inequality[2] we have $Pr(|Y - \mu_Y| < \sigma_Y \lambda) > 1 - \frac{1}{\lambda^2}$. Let Y' be the random variable given by Y conditioned on being in the interval $J = [\mu_Y - \sigma_Y\lambda, \mu_Y + \sigma_Y\lambda] \cap \mathbb{Z}$. Note $|J| \leq 2\sigma_Y\lambda + 1$. Then by lemma (3.2) there exists an integer $k \in [i_1 - (\mu_Y + \sigma_Y\lambda), i_2 - (\mu_Y - \sigma_Y\lambda)]$ s.t. $\mathbb{P}\big(\tau_k(Y') \in S\big) \geq \frac{|S|}{|I|+|J|-1} = \frac{d}{1+\frac{2\sigma_Y\lambda}{|I|}}$. Therefore we have $\mathbb{P}\big(\tau_k(Y) \in S\big) \geq \mathbb{P}(Y \in J)\mathbb{P}\big(\tau_k(Y') \in S\big) \geq (1 - \frac{1}{\lambda^2})(\frac{d}{1+\frac{2\sigma_Y\lambda}{|I|}})$. □

It is possible to construct sets of intervals which under the image of H act as shifts of each other. For example in base-10 (and assuming that $H(0) = 0, H(1) = 1$) if the random variable X_1 is induced by $[1100, 1199]$ and X_2 induced by $[0, 99]$ then $H(X_1) = H(X_2) + 2$.

We will now further expand on the example above. Consider $n \in \mathbb{N}$, s.t. $4 \mid n$. Let $B_0 = [0, b^{\frac{3n}{4}} - 1]$ and for $k = 1, \ldots, \frac{n}{4}$ consider the interval

$$B_k = [b^{n-1} + b^{n-2} + \cdots + b^{n-k}, b^{n-1} + b^{n-2} + \cdots + b^{n-k} + b^{\frac{3n}{4}} - 1]$$

[2] We certainly could do better than Chebyshev's Inequality here. However, the bounds it gives will suit our purposes fine.

Then the B_k will all be n-strict (with exception of B_0), and a random integer $x \in B_k$ will have the following base-b expansion:

$$x = \underbrace{11\ldots1}_{\text{k digits}}\ \underbrace{00\ldots0}_{\frac{n}{4}-k \text{ digits}}\ \underbrace{X_iX_{i-1}\ldots X_1}_{\frac{3n}{4}\text{ digits}}$$

That is x will have its first k digits $= 1$, the next $\frac{n}{4} - k$ digits $= 0$, and the remaining $\frac{3n}{4}$ digits will be iid random variables X_i taking values $0, 1, \ldots, b-1$ with uniform distribution.

Let Y_0 be the random variable induced by B_0, and Y_k be induced by B_k. Again because $H(0) = 0, H(1) = 1$,

$$H(Y_k) = H(Y_0) + k = \tau_k\left(H(Y_0)\right)$$

Recall $H(Y_0)$ has mean $\frac{3n}{4}\mu$, and variance $\frac{3n}{4}\sigma^2$. Consider an interval $I = [i_1, i_2]$ containing a set of type-C integers S, and let $\lambda > 0$. By lemma (3.3)

$$\exists k' \in \Big[i_1 - \Big(\frac{3n}{4}\mu + \sqrt{\frac{3n}{4}}\lambda\sigma\Big), i_2 - \Big(\frac{3n}{4}\mu - \sqrt{\frac{3n}{4}}\lambda\sigma\Big)\Big] \text{ s.t.}$$

$$\mathbb{P}\Big(\tau_{k'}\left(H(Y_0)\right) \in S\Big) \geq \Big(1 - \frac{1}{\lambda^2}\Big)\Big(\frac{d}{1 + \frac{\sqrt{3n}\lambda\sigma}{|I|}}\Big) \tag{6}$$

Thus if $I \subseteq \Big[1 + \frac{3}{4}n\mu + \lambda\sigma\sqrt{\frac{3}{4}n}, \frac{1}{4}n + \frac{3}{4}n\mu - \lambda\sigma\sqrt{\frac{3}{4}n}\Big]$ then $1 \leq k' \leq \frac{n}{4}$. Setting $k = k'$ produces the interval B_k, which will be n-strict with type-C density $\geq (6)$. In fact we have proven the following

Theorem 3.4. *Let $n \in \mathbb{N}$ with $4 \mid n$, let $C \subset \mathbb{Z}^+$. Given $\lambda > 0$ let $J_{n,\lambda} = \Big[1 + \frac{3}{4}n\mu + \lambda\sigma\sqrt{\frac{3}{4}n}, \frac{1}{4}n + \frac{3}{4}n\mu - \lambda\sigma\sqrt{\frac{3}{4}n}\Big]$. Suppose $\exists I \subseteq J_{n,\lambda}$ with type-C density $= d$. Then there exists an n-strict interval I_2 s.t. I_2 has type-C density $\geq \Big(1 - \frac{1}{\lambda^2}\Big)\Big(\frac{d}{1+\frac{\sqrt{3n}\sigma\lambda}{|I|}}\Big)$.*

Suppose we have found an n-strict interval I with a desired type-C density d. We can then apply the previous theorem in order to construct a second interval I_2 which is n_2-strict for some n_2 much larger than n assuming that $\exists n_2$ s.t. $4 \mid n_2$ and $I \subseteq J_{n_2,\lambda}$. The question now becomes, how large must n be in order to guarantee the existence of n_2 and λ (λ large enough as to not lose much density) s.t. $I \subseteq J_{n_2,\lambda}$.

Such an n_2 if it exists would be $\approx \frac{4}{3\mu}b^n$ (looking at the dominant term of the left endpoint of $J_{n_2,\lambda}$), which would give

$$|J_{n_2,\lambda}| \approx \frac{b^n}{3\mu} - \frac{\lambda\sigma b^{\frac{n}{2}}}{\sqrt{\mu}}$$

If this n_2 is to work we'd surely need $b^{\frac{3n}{4}} = |I| < |J_{n_2,\lambda}|$. If λ is small then it is quite easy to find such an n_2, but we may in turn lose too much density. However

as n gets large we can find such an n_2 while keeping λ large. In an attempt to make calculations as simple as possible, the next theorem sets $\lambda = b^{\frac{n}{8}}$ (this is chosen for simplicity, not to optimize the result) and then finds how large n must be in order for an appropriate n_2 to exist. The proof follows from a number of routine calculations and estimations, some of which we have left for the appendix.

Theorem 3.5. *Suppose I is n-strict with type-C density $= d$, where n satisfies the bound ($\boldsymbol{B}$). Then $\exists n_2 \geq \frac{b^{n-1}}{\mu}$ and an n_2-strict interval I_2 s.t. the type-C density of $I_2 \geq d\left(1 - b^{\frac{-n}{4}}\right)\left(1 - \frac{2\sigma}{\sqrt{\mu}} b^{\frac{-n}{8}}\right)$.*

Proof. As before, let $J_{m,\lambda} = \left[1 + \frac{3}{4}m\mu + \lambda\sigma\sqrt{\frac{3}{4}m}, \frac{1}{4}m + \frac{3}{4}m\mu - \lambda\sigma\sqrt{\frac{3}{4}m}\right]$. We assumed that I is n-strict, so $|I| = b^{\frac{3n}{4}}$. Call the left endpoint of I, a, and write $I = [a, a + b^{\frac{3}{4}n} - 1]$. Setting $\lambda = b^{\frac{n}{8}}$ we attempt to find an n_2 divisible by 4 s.t. $I \subseteq J_{n_2,\lambda}$. It would be prudent to consider $f(m) = 1 + \frac{3}{4}m\mu + \lambda\sigma\sqrt{\frac{3}{4}m}$ which represents the left endpoint of $J_{m,\lambda}$. We first find an integer n_2 s.t. $f(n_2)$ is close to yet still $\leq a$. By lemma (6.2) in the appendix, assuming n satisfies (**B**) it follows that there exists n_2 s.t.

- $4 \mid n_2$
- $\frac{b^{n-1}}{\mu} \leq n_2 \leq \frac{4}{3\mu} b^n$
- $0 \leq a - f(n_2) \leq 3\mu + 1$

We now check that $I \subseteq J_{n_2}$ in order to invoke Theorem 3.4. We already have that the left endpoint $f(n_2) \leq a$. So now check the right endpoints of I and J_{n_2}. We need

$$a - 1 + b^{\frac{3n}{4}} \leq \frac{n_2}{4} + \frac{3}{4}n_2\mu - \lambda\sigma\sqrt{\frac{3}{4}n_2} \tag{7}$$

The above is equivalent to

$$a - \left(\frac{3}{4}n_2\mu + \lambda\sigma\sqrt{\frac{3}{4}n_2} + 1\right) + b^{\frac{3n}{4}} \leq \frac{n_2}{4} - 2\lambda\sigma\sqrt{\frac{3}{4}n_2}$$

Simplifying, the above follows from showing

$$a - f(n_2) + b^{\frac{3n}{4}} + 2\lambda\sigma\sqrt{\frac{3}{4}n_2} \leq \frac{n_2}{4}$$

Now letting

$$LHS = a - f(n_2) + b^{\frac{3}{4}n} + 2\lambda\sigma\sqrt{\frac{3}{4}n_2}$$

Then

$$LHS \leq 3\mu + 1 + b^{\frac{3}{4}n} + \lambda\sigma\sqrt{3n_2}$$

Using the fact that $\lambda = b^{\frac{n}{8}}, n_2 \leq \frac{4b^n}{3\mu}$ we get:

$$LHS \leq 3\mu + 1 + b^{\frac{3}{4}n} + 2\frac{\sigma}{\sqrt{\mu}} b^{\frac{5n}{8}}$$

And letting $RHS = \frac{n_2}{4}$, by assumptions on n_2

$$RHS \geq \frac{b^n}{4b\mu}$$

So (7) follows from

$$3\mu + 1 + b^{\frac{3}{4}n} + 2\frac{\sigma}{\sqrt{\mu}} b^{\frac{5n}{8}} \leq \frac{b^n}{4b\mu}$$

The above is exactly the bound (**B3**). Therefore $I \subseteq J_{n_2,\lambda}$. Thus by applying Theorem (3.4) with $\lambda = b^{\frac{n}{8}}$, $\exists$ an n_2-strict interval I_2 s.t.

$$\text{type-C density of } I_2 \geq \left(1 - \frac{1}{b^{\frac{n}{4}}}\right)\left(\frac{d}{1 + \frac{\sqrt{3n_2}\sigma b^{\frac{n}{8}}}{b^{\frac{3n}{4}}}}\right)$$

Since $n_2 \leq \frac{4}{3\mu} b^n$ it follows that,

$$\frac{1}{1 + \frac{\sqrt{3n_2}\sigma b^{\frac{n}{8}}}{b^{\frac{3n}{4}}}} \geq \frac{1}{1 + \frac{2\sigma}{\sqrt{\mu}} b^{\frac{-n}{8}}} \geq 1 - \frac{2\sigma}{\sqrt{\mu}} b^{\frac{-n}{8}}$$

Thus we conclude the type-C density of $I_2 \geq d\left(1 - b^{\frac{-n}{4}}\right)\left(1 - \frac{2\sigma}{\sqrt{\mu}} b^{\frac{-n}{8}}\right)$. □

Apply the previous theorem to our starting n_1-strict interval I_1 to get an n_2-strict interval I_2. Since $n_2 > n_1$, we can apply theorem (3.5) again on I_2. Continuing in this fashion produces a sequence of integers (n_i), and n_i-strict intervals (I_i) each with type-C density d_i s.t.

- $n_{i+1} \geq \frac{b^{n_i - 1}}{\mu}$
- $d_{i+1} \geq d_i \left(1 - b^{\frac{-n_i}{4}}\right)\left(1 - \frac{2\sigma}{\sqrt{\mu}} b^{\frac{-n_i}{8}}\right)$

The second condition implies

$$\forall i, d_i \geq d_1 \prod_{i=1}^{\infty} \left((1 - b^{\frac{-n_i}{4}})(1 - \frac{2\sigma}{\sqrt{\mu}} b^{\frac{-n_i}{8}})\right) \tag{8}$$

Looking at the first condition, the sequence (n_i) exhibits super exponential growth which clearly implies that $\forall i, n_i \geq i n_1$. The only worry would be when μ is very large, but the bound (**B**) takes care of this case.

Note for $x, \alpha \in \mathbb{R}^+$, if $x \geq 2\alpha > 0$ then

$$1 - \alpha x^{-1} \geq \frac{1}{1 + 2\alpha x^{-1}} \geq e^{-2\alpha x^{-1}}$$

Applying this first with $\alpha = 1, x = b^{\frac{n_i}{4}}$ and second with $\alpha = \frac{2\sigma}{\sqrt{\mu}}, x = b^{\frac{n_i}{8}}$, we may now rewrite (8) as

$$d_i \geq d_1 \exp\left(\sum_{i=1}^{\infty} -2b^{\frac{-n_i}{4}} - \frac{4\sigma}{\sqrt{\mu}} b^{\frac{-n_i}{8}}\right)$$

Using the fact that $\forall i, n_i \geq in_1$, the sum in the previous inequality will become the sum of 2 geometric series, one with $r = b^{\frac{-n_1}{4}}, a = -2b^{\frac{-n_1}{4}}$, and the second with $r = b^{\frac{-n_1}{8}}, a = \frac{-4\sigma}{\sqrt{\mu}} b^{\frac{-n_1}{8}}$. Recall that a geometric series with $|r| < 1$, and first term a sums to

$$\frac{a}{1-r}$$

Therefore the first series sums to $\frac{-2b^{\frac{-n_1}{4}}}{1-b^{\frac{-n_1}{4}}}$, the second sums to $\frac{-4\sigma b^{\frac{-n_1}{8}}}{\sqrt{\mu}(1-b^{\frac{-n_1}{8}})}$. Finally after simplifying we conclude that

$$\forall i, d_i \geq d_1 \exp\left(\frac{-2}{b^{\frac{n_1}{4}} - 1} + \frac{-4\sigma}{\sqrt{\mu}(b^{\frac{n_1}{8}} - 1)}\right)$$

Thus we have proven the following

Theorem 3.6. *Assume $\exists n_1$ satisfying the bound (**B**), and an n_1-strict interval I_1 with type-C density $= d_1$. Then $\forall N \in \mathbb{N}, \exists n > N$, and n-strict interval I s.t.*

$$\text{the type-C density of } I \geq d_1 \, exp\left(\frac{2}{1 - b^{\frac{n_1}{4}}} + \frac{4\sigma}{\sqrt{\mu}(1 - b^{\frac{n_1}{8}})}\right)$$

4. Main Result

As in the previous section we continue to assume H is a generalized b-happy function with digit mean and variance μ, σ^2. Also we assume that we have experimentally found a suitable starting n_1-strict interval I_1 with type-C density d_1 for some $C \subset \mathbb{N}$. As in section 2, for positive integer m, let Y_m be the random variable induced by the interval $[0, b^m - 1]$.

In this section we give a proof of the following

Theorem 4.1. *Suppose I_1 is n_1-strict, where n_1 satisfies the bound (**B**) and I_1 has type-C density $= d_1$. Then the upper density of the set of type-C integers is $\geq d_1 \, exp\left(\frac{2}{1-b^{\frac{n_1}{4}}} + \frac{4\sigma}{\sqrt{\mu}(1-b^{\frac{n_1}{8}})}\right)$*

The digit mean and digit variance for the case $(e, b) = (2, 10)$ is 28.5 and 721.05 respectively. In this case, if $n > 13$ then it satisfies bound (**B**). After performing a computer search we find that the density of happy numbers in the interval $[10^{403}, 10^{404} - 1]$ is $> .185773$, thus there exists a 404-strict interval having at least this density of happy numbers as a subset (it is necessary here

that $4 \mid 404$). Letting $\delta(n) = \left(\frac{2}{1-b^{\frac{n}{4}}} + \frac{4\sigma}{\sqrt{\mu}(1-b^{\frac{n}{8}})}\right)$ we find that $|\delta(404)| < 10^{-49}$. For the lower density, the type-$\{1\}$ density of $[10^{2367}, 10^{2368}-1]$ is $< .11379$. This implies that there is a 2368-strict interval with type-$\{4\}$ density $> 1 - .11379$, apply the main result to conclude that the upper density of type-$\{4\}$ integers is $> 1 - .1138$. This gives the following

Corollary 4.2. *Let $\underline{d}$ and $\bar{d}$ be the lower and upper density of $(2, 10)$-happy numbers respectively. Then $\underline{d} < .1138, \bar{d} > .18577$.*

In general, by taking C' to be the union of all cycles other than C. We get as a corollary an upper bound on the lower density of type-C integers.

The proof of theorem (4.1) is technical despite having a rather intuitive motivation. For the sake of clarity we first give a sketch of how to use theorem (3.6) with theorem (2.7) in order to prove a lower bound on the upper density of type-C numbers.

Given our starting n_1-strict interval I_1 apply the theorem (3.6) to construct an n-strict interval I (n large) with type-C density $\geq (1 - o(1))d_1$. Pick m_1 s.t. the mean of $H(Y_{m_1}) = \mu m_1$ lands in the interval I. Since $I \subseteq [b^{n-1}, b^n]$ we have $m_1 = O(b^n)$. This implies the standard deviation is $O(b^{\frac{n}{2}})$. This will be much less than $|I| = b^{\frac{3n}{4}}$.

Next, use lemma (3.3) with a large λ, to find a shift $\tau_k\big(H(Y_{m_1})\big)$ which has probability $\geq (1 - o(1))d_1$ of being type-C (note this $k \leq |I| = b^{\frac{3n}{4}}$). The mean of $\tau_k\big(H(Y_{m_1})\big) = \mu m_1 + k$. Clearly there exists an integer m_2 s.t.

$$|\mu m_2 - (\mu m_1 + k)| \leq \mu$$

The means of $H(Y_{m_2})$ and $\tau_k\big(H(Y_{m_1}\big)$ are almost equal. Since k is much smaller than m_1, m_2 we apply theorem (2.7) to argue that within some (large) neighbourhood of the means, the distributions of $H(Y_{m_2})$ and $\tau_k\big(H(Y_{m_1})\big)$ are nearly identical. We conclude that $\mathbb{P}\big(H(Y_{m_2}) \text{ is type-C}\big) \geq (1 - o(1))d_1$. Doing this for (arbitrarily) large n (and therefore m_2) we find interval $[0, b^{m_2} - 1]$ with type-C density $\geq d_1\,(1 - o(1))$. This in turn shows that the upper density of type-C numbers is $\geq d_1\,(1 - o(1))$.

4.1. Some Lemmas

Making the above sketch fully rigorous requires the following ϵ argument. We have broken some of the pieces down for 3 lemmas. The proofs primarily consist of calculations and we leave them for after the proof of the main result. We first make the following remarks about lemmas (4.3) and (4.5).

- It is not important how large the required N must be.
- Lemma (4.5) part (1) is the only place where theorem (2.7) is needed.

Lemma 4.3. *$\exists N$ sufficiently large s.t. $\forall n > N, \forall I$ n-strict, $\exists m \in \mathbb{N}$ s.t.*

$$[\mu m - \sigma m^{\frac{5}{8}}, \mu m + \sigma m^{\frac{5}{8}}] \subseteq I$$

Lemma 4.4. *Let $\epsilon > 0$ be given (assume as well that $\epsilon \leq 1$). Let $\lambda = \sqrt{\frac{6}{\epsilon}}$. Then $\exists N(\epsilon)$ s.t. if any n, m_1, I satisfy:*

- $n > N$
- $m_1 \in [\frac{b^{n-1}}{\mu}, \frac{b^n}{\mu}]$
- *I is n-strict*

Then the following hold:

1. $\lambda \leq m_1{}^{\frac{1}{8}}$
2. $|1 - \frac{1}{1+\frac{2\lambda\sigma\sqrt{m_1}}{b^{\frac{3n}{4}}}}| \leq \frac{\epsilon}{6}$

Lemma 4.5. *Let $\epsilon > 0$ be given (assume as well that $\epsilon \leq 1$). Let $T = \frac{2\sqrt{6}}{\sqrt{\epsilon}}$, and $\lambda = \frac{\sqrt{6}}{\sqrt{\epsilon}}$. Then $\exists N(\epsilon)$ s.t. if any n, m_1, m_2, k, I satisfy:*

- $n > N$
- $m_1 \in [\frac{b^{n-1}}{\mu}, \frac{b^n}{\mu}], m_2 \in [\frac{b^{n-2}}{\mu}, \frac{b^{n+1}}{\mu}]$
- $|k| \leq b^{\frac{3n}{4}}$
- $|\mu m_1 + k - \mu m_2| \leq \mu$
- *I is n-strict*

Then the following hold:

1. *For $i = 1, 2$:* $\max_{|t| \leq T} |1 - \frac{Pr(H(X_{m_i}) = \lfloor \mu m_i + t\sigma\sqrt{m_i} \rfloor)}{\frac{e^{\frac{t^2}{2}}}{\sqrt{2\pi m_i}\sigma}}| \leq \frac{\epsilon}{6}$
2. $|1 - \sqrt{\frac{m_1}{m_2}}| \leq \frac{\epsilon}{6}$
3. *If $\exists t_1, t_2$ s.t. $t_1 \in [\frac{-T}{2}, \frac{T}{2}]$ and $\mu m_1 + k + t_1\sigma\sqrt{m_1} = \mu m_2 + t_2\sigma\sqrt{m_2}$. Then $t_2 \in [-T, T]$ and $|1 - e^{\frac{t_2{}^2 - t_1{}^2}{2}}| \leq \frac{\epsilon}{6}$.*

4.2. Proof of Theorem 4.1

Proof. Let $f(n) = |\{i \in \mathbb{Z}^+ | i \leq n, i \text{ is type-C}\}|$. We must show that

$$\limsup \frac{f(n)}{n} \geq d_1 \exp\left(\frac{2}{1 - b^{\frac{n_1}{4}}} + \frac{4\sigma}{\sqrt{\mu}(1 - b^{\frac{n_1}{8}})}\right)$$

It suffices to show $\forall \epsilon > 0, \forall N_1 \in \mathbb{N}, \exists m > N_1$ s.t. the interval $[0, b^m - 1]$ has type-C density $\geq d_1 \exp\left(\frac{2}{1-b^{\frac{n}{4}}} + \frac{4\sigma}{\sqrt{\mu}(1-b^{\frac{n}{8}})}\right)(1 - \epsilon)$. Let ϵ and N_1 be arbitrary (assume only that $\epsilon \leq 1$). Set $T = \frac{2\sqrt{6}}{\sqrt{\epsilon}}$ (we will be applying the local limit

law for the bounded interval $[-T,T]$ later in the proof). Also in anticipation of applying lemma (3.3) set $\lambda = \frac{\sqrt{6}}{\sqrt{\epsilon}}$.

First pick $N > N_1$ large enough to apply lemmas (4.3), (4.4), and (4.5). Invoke theorem (3.6) to construct an n-strict interval I for $n > N$ with type-C density d. Then

$$d \geq d_1 \exp\left(\frac{2}{1-b^{\frac{n_1}{4}}} + \frac{4\sigma}{\sqrt{\mu}(1-b^{\frac{n_1}{8}})}\right) \tag{9}$$

For $m \in \mathbb{N}$ let Y_m be the random variable induced by the interval $[0, b^m - 1]$, also let

$$J_m = [\mu m - \sigma m^{\frac{5}{8}}, \mu m + \sigma m^{\frac{5}{8}}]$$

Recall again that

$$\mathbf{E}[H(Y_m)] = \mu m$$

$$\mathbf{Var}[H(Y_m)] = \sigma^2 m$$

Hence J_m is where the "bulk" of the distribution of $H(Y_m)$ lands. Pick m_1 s.t. $J_{m_1} \subseteq I$ (the existence of which is guaranteed by lemma (4.3)). Note that $m_1 \in [\frac{b^{n-1}}{\mu}, \frac{b^n}{\mu}]$ since I is n-strict. Let S be the set of type-C integers in I. Apply lemma (3.3) on $H(Y_{m_1})$ and λ to find integer k s.t.

$$\mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) \in S\Big) \geq d\left(1 - \frac{1}{\lambda^2}\right)\left(\frac{1}{1 + \frac{2\lambda\sigma\sqrt{m_1}}{|I|}}\right) \tag{10}$$

Since $J_{m_1} \subseteq I$ and $|I| = b^{\frac{3n}{4}}$ it follows that $k \leq b^{\frac{3n}{4}}$.

Recall the proof of lemma (3.3). In particular we ignored the tails of the distribution of $H(Y_{m_1})$ outside of $\lambda\sigma\sqrt{m_1}$ from the mean. Notice as well that $\lambda \leq m_1{}^{\frac{1}{8}}$ (by lemma (4.4) part 1). Calling S' the set of type-C integers in the shifted interval $\tau_k(J_{m_1})$, we may replace (10) by the stronger conclusion

$$\sum_{i \in S'} \mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big) \geq d\Big(1 - \frac{1}{\lambda^2}\Big)\Big(\frac{1}{1 + \frac{2\lambda\sigma\sqrt{m_1}}{|I|}}\Big)$$

Using the assumption that $\lambda = \sqrt{\frac{6}{\epsilon}}$ and part 2 of lemma (4.4) we simplify the above as

$$\sum_{i \in S'} \mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big) \geq d(1 - \frac{\epsilon}{6})^2 \tag{11}$$

Now pick $m_2 \in \mathbb{N}$ s.t. $|m_1\mu + k - m_2\mu| \leq \mu$. Since $|k| \leq b^{\frac{3n}{4}}$, it follows that $m_2 \in [\frac{b^{n-2}}{\mu}, \frac{b^{n+1}}{\mu}]$. In particular m_1, m_2, n, k, I now all satisfy the conditions of lemma (4.5). Let Y_{m_2} be the r.v. induced by the interval $[0, b^{m_2} - 1]$. It remains to show that near the mean of $\tau_k\big(H(X_{m_1})\big)$, the distributions of $\tau_k\big(H(Y_{m_1})\big)$ and $H(Y_{m_2})$ are similar, this will imply that the interval $[0, b^{m_2} - 1]$ contains a large amount of type-C integers. Making this precise, we prove the following

Claim 1. *For integers $i \in \tau_k\big(J_{m_1}\big)$*

$$\frac{\mathbb{P}\big(H(Y_{m_2}) = i\big)}{\mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big)} \geq (1 - \frac{\epsilon}{6})^4$$

Proof. We let $i \in J_{m_1}$ be fixed and pick t_1, t_2 s.t.

$$i = \mu m_1 + k + t_1 \sigma \sqrt{m_1} = \mu m_2 + t_2 \sigma \sqrt{m_2}$$

It is important now that we had chosen $\lambda = \frac{T}{2}$, this implies $|t_2| \leq T$ (see lemma (4.5) part (3)). We can use the local limit law to estimate the distributions of $\tau_k\big(H(Y_{m_1})\big)$ and $H(Y_{m_2})$. By lemma (4.5) part (1)

$$\mathbb{P}\big(H(Y_{m_2}) = i\big) = \mathbb{P}\big(H(Y_{m_2}) = \mu m_2 + t_2 \sigma \sqrt{m_2}\big) \geq \frac{e^{\frac{t_2{}^2}{2}}}{2\pi\sigma\sqrt{m_2}}(1 - \frac{\epsilon}{6})$$

and

$$\mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big) = \mathbb{P}\big(H(Y_{m_1}) = \mu m_1 + t_1 \sigma \sqrt{m_1}\big) \leq \frac{e^{\frac{t_1{}^2}{2}}}{2\pi\sigma\sqrt{m_1}}(1 + \frac{\epsilon}{6})$$

Hence

$$\frac{\mathbb{P}\big(H(Y_{m_2}) = i\big)}{\mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big)} \geq e^{\frac{t_2{}^2 - t_1{}^2}{2}} \frac{\sqrt{m_1}}{\sqrt{m_2}} \frac{1 - \frac{\epsilon}{6}}{1 + \frac{\epsilon}{6}}$$

Which by lemma (4.5) parts (2),(3)

$$\geq (1 - \frac{\epsilon}{6})^4$$

□

Let d_2 be the density of type-C numbers in the interval $[0, b^{m_2} - 1]$. Then

$$d_2 \geq \sum_{i \in S'} \mathbb{P}\big(H(Y_{m_2}) = i\big) = \sum_{i \in S'} \frac{\mathbb{P}\big(H(Y_{m_2}) = i\big)}{\mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big)} \mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big)$$

Which by claim 1

$$\geq (1 - \frac{\epsilon}{6})^4 \sum_{i \in S'} \mathbb{P}\Big(\tau_k\big(H(Y_{m_1})\big) = i\Big)$$

And by equation (11):

$$\geq d(1 - \frac{\epsilon}{6})^6 \geq d(1 - \epsilon)$$

Finally by equation (9):

$$d_2 \geq d_1 \exp\left(\frac{2}{1 - b^{\frac{n_1}{4}}} + \frac{4\sigma}{\sqrt{\mu}(1 - b^{\frac{n_1}{8}})}\right)(1 - \epsilon)$$

This concludes the proof.

□

4.3. Proof of Lemma 4.2

Proof. Let $J_m = [\mu m - \sigma m^{\frac{5}{8}}, \mu m + \sigma m^{\frac{5}{8}}]$. Then $|J_m| \leq 2\sigma m^{\frac{5}{8}} + 1$. We assumed I is n-strict, so $I \subseteq [b^{n-1}, b^n - 1]$, let a be the left endpoint of I. Picking m s.t. $\mu m \in I$ will give $m = O(b^n)$, which in turn gives $|J_m| = O(b^{\frac{5n}{8}}) \ll |I| = b^{\frac{3n}{4}}$. Comparing the growth rates of $|J_m|$ and $|I|$ it is clear that we can pick N_1 large enough s.t. $n > N_1 \implies \exists m$ s.t. $J_m \subseteq I$. □

4.4. Proof of Lemma 4.3

Proof. We find N_1, N_2 for the 2 parts respectively, then choose $N = \max(N_1, N_2)$.

1. λ is a fixed constant here and it is assumed that $m_1 \geq \frac{b^{n-1}}{\mu}$, so the result is trivial (this gives N_1).

2. For $x > 0$ to show $|1 - \frac{1}{1+x}| \leq \frac{\epsilon}{6}$, it is equivalent to show

$$1 - \frac{\epsilon}{6} \leq \frac{1}{1+x} \leq 1 + \frac{\epsilon}{6}$$

Or equivalently

$$(1 - \frac{\epsilon}{6})(1+x) \leq 1 \leq (1 + \frac{\epsilon}{6})(1+x)$$

The above follows if $x \leq \frac{\epsilon}{6}$. So we need N large enough s.t. $\frac{2\sigma\lambda\sqrt{m_1}}{b^{\frac{3n}{4}}} \leq \frac{\epsilon}{6}$. Using the assumption that $m_1 \leq \frac{b^n}{\mu}$ we get

$$\frac{2\sigma\lambda\sqrt{m_1}}{b^{\frac{3n}{4}}} \leq \frac{2\sigma\lambda}{\sqrt{\mu} b^{\frac{n}{4}}}$$

And thus the result follows if

$$\frac{12\sigma\lambda}{\sqrt{\mu}\epsilon} \leq b^{\frac{n}{4}}$$

Hence picking $N_2 \geq 4\log_b(\frac{12\sigma\lambda}{\sqrt{\mu}\epsilon})$ suffices.

□

4.5. Proof of Lemma 4.4

Proof. We find N_1, N_2, N_3, for the 3 parts respectively, and then choose $N = \max(N_1, N_2, N_3)$.

1. As shown in section 2, $H(Y_{m_1})$ is the sum of m_1 iid random variables each with mean μ and variance σ^2. Thus the result follows from applying theorem (2.7) to the sum $\sum_{i=1}^{\infty} H(X_{m_i})$ with finite interval $[-T, T]$, getting M_1 large enough s.t. for $m_1 > M_1$, for all $t \in [-T, T]$, the fuzzy term $O(m_1{}^{\frac{-1}{2}}) \leq \frac{\epsilon}{6}$. Do the same on the sum for $H(X_{m_2})$ to get M_2. By the assumption $m_1, m_2 \geq \frac{b^{n-1}}{\mu}$ choosing $N_3 = \max(\mu M_1 + 1, \mu M_2 + 1)$ suffices.

2. We can ignore the square root, it suffices to show that

$$|1 - \frac{m_1}{m_2}| \leq \frac{\epsilon}{6} \tag{12}$$

By assumption

$$|\mu m_1 + k - \mu m_2| \leq \mu$$

Divide through by μm_2 and get

$$|1 - \frac{m_1}{m_2} - \frac{k}{\mu m_2}| \leq \frac{1}{m_2}$$

This follows from

$$\frac{k}{\mu m_2} - \frac{1}{m_2} \leq 1 - \frac{m_1}{m_2} \leq \frac{1}{m_2} + \frac{k}{\mu m_2}$$

Thus (12) follows from showing $\frac{1}{m_2} + \frac{k}{\mu m_2} \leq \frac{\epsilon}{6}$. Using the assumption $m_2 \geq \frac{b^{n-2}}{\mu}$ and $|k| \leq b^{\frac{3n}{4}}$ it follows that

$$\frac{1}{m_2} + \frac{k}{\mu m_2} \leq \mu b^{-(n-2)} + b^{\frac{-n}{4}+2}$$

Picking $N_4 = \max(\log_b(\frac{12\mu}{\epsilon}) + 2, 4\log_b(\frac{12}{\epsilon}) + 2)$ works.

3. We first find N' s.t. $n > N' \implies t_2 \in [-T, T]$. Starting with the assumption that

$$\mu m_1 + k + t_1\sigma\sqrt{m_1} = \mu m_2 + t_2\sigma\sqrt{m_2}$$

Rearranging it, using the facts that $|\mu m_1 + k - \mu m_2| \leq \mu$ and $|t_1| \leq \frac{T}{2}$ this implies that

$$|t_2| \leq \frac{\mu}{\sigma\sqrt{m_2}} + \frac{T\sqrt{m_1}}{2\sqrt{m_2}}$$

We assumed that $m_2 \geq \frac{b^{n-2}}{\mu}$, also in part (2) we showed $\exists N_4$ s.t. $n > N_4 \implies \frac{\sqrt{m_1}}{\sqrt{m_2}} \leq (1 + \frac{\epsilon}{6}) \leq \frac{7}{6}$. Hence

$$|t_2| \leq \frac{\mu^2}{\sigma b^{\frac{n-2}{2}}} + \frac{7T}{12}$$

Pick $N' > N_4$ large enough s.t. $n > N' \implies \frac{\mu^2}{\sigma b^{\frac{n-2}{2}}} \leq \frac{5T}{12}$. This will take care of the size of t_2.

Now we must show that $\exists N''$ large enough s.t. $n > N'' \implies$

$$|1 - e^{\frac{{t_2}^2 - {t_1}^2}{2}}| \leq \frac{\epsilon}{6}$$

After finding N'' we set $N_5 = \max(N', N'')$. Note that for $x \in \mathbb{R}$ to show $|1 - e^x| \leq \frac{\epsilon}{6}$ it suffices to show that

$$\ln(1 - \frac{\epsilon}{6}) \leq x \leq \ln(1 + \frac{\epsilon}{6})$$

Thus letting $\epsilon^* = \min(\ln(1+\frac{\epsilon}{6}), |\ln(1-\frac{\epsilon}{6})|)$ we find N'' s.t. $n > N'' \implies$

$$|\frac{{t_2}^2 - {t_1}^2}{2}| \leq \epsilon^*$$

It was assumed that

$$\mu m_1 + k + t_1\sigma\sqrt{m_1} = \mu m_2 + t_2\sigma\sqrt{m_2}$$

Or equivalently

$$\mu m_1 + k - \mu m_2 = t_2\sigma\sqrt{m_2} - t_1\sigma\sqrt{m_1}$$

Applying the assumption that $|LHS| \leq \mu$ and dividing both sides by $\sigma\sqrt{m_2}$ we get

$$|t_2 - t_1\sqrt{\frac{m_1}{m_2}}| \leq \frac{\mu}{\sqrt{m_2}\sigma} \iff$$

$$|t_2 - t_1 + t_1(1-\sqrt{\frac{m_1}{m_2}})| \leq \frac{\mu}{\sqrt{m_2}\sigma} \implies$$

$$|t_2 - t_1| \leq \frac{\mu}{\sqrt{m_2}\sigma} + \left|t_1(1-\sqrt{\frac{m_1}{m_2}})\right|$$

Recall by assumption $m_2 \geq \frac{b^(n-2)}{\mu}$, $|t_1|, |t_2| \leq T$, and by the proof of part (2) $\left|1-\sqrt{\frac{m_1}{m_2}}\right| \leq \mu b^{-(n-2)} + b^{\frac{-(n-2)}{4}}$. Putting this together it follows that

$$\left|\frac{{t_2}^2 - {t_1}^2}{2}\right| = \left|(\frac{t_2+t_1}{2})(t_2-t_1)\right| \leq T\Big(\frac{\mu^{\frac{3}{2}} b^{\frac{-(n-2)}{2}}}{\sigma} + T(\mu b^{-(n-2)} + b^{\frac{-(n-2)}{4}})\Big)$$

Now since T, μ, σ, b are all constants, $\lim_{n\to\infty} RHS = 0$ so $\exists N''$ s.t. $n > N'' \implies RHS \leq \epsilon^*$. Finally set $N_5 = \max(N', N'')$. $\square$

5. Experimental Data

The data[3] presented in this section is the results of short computer searches, so the bounds surely can be improved greatly with more computing time. Floating point approximation with conservative rounding was used.

5.1. Finding an Appropriate n-strict Interval

If n is divisible by 4, and the interval $[b^{n-1}, b^n - 1]$ has type-C density d, the there exists an n-strict interval with type-C density $>= d$, which we may apply theorem (4.1) to. The density of $[b^{n-1}, b^n - 1]$ for various n can be quickly calculated by first computing the densities of intervals of the form $[0, b^n - 1]$, the algorithm for which was discussed in section (2.2). After an appropriate n-strict interval is found, we check to see that n satisfies the bound (**B**), compute the error term, and find the desired bound. Our results show that in almost all cases, the asymptotic density does not exist.

[3]Data generated by fellow graduate student, Patrick Devlin. Also it was double checked for smaller n using a slower Maple program written independently by the author.

5.2. Explanation of Results

The following information is given in tables (in the order of column in which they appear):

1. The cycle in which type-C densities are being computed.
2. The lower bound on the upper density (UD) implied by theorem (4.1)
3. The upper bound on the lower density (LD) implied by theorem (4.1)
4. The n s.t. the interval $[b^{n-1}, b^n - 1]$ is used to find the bound (denoted as UD n or LD n
5. The $\delta(n) = \left(\frac{2}{1-b^{\frac{n}{4}}} + \frac{4\sigma}{\sqrt{\mu}(1-b^{\frac{n}{8}})}\right)$ part of the error term for Theorem (4.1) (we only present an upper bound on $|\delta(n)|$, the true number is always negative). In all cases the error is small enough not to affect the bounds as we only give precision of about 5 or 6 decimal places.

5.2.1. Cubing the digits

In this case if $n > 16$ then it satisfies bound (**B**). Table 1 shows the results for the cycles when studying the $(3, 10)$-happy function. There are 9 possible cycles. Figure 2 graphs the density of type-$\{1\}$ integers less than 10^n. It is easy to prove in this case that $3 \mid n \iff n$ is type-$\{153\}$.

Table 1: Bounds for the Cycles Appearing for Digit Sequence $[0, 1, 8, 27, 64, 125, 216, 343, 512, 729]$

Cycle	UD	LD	UD n	LD n	UD $\delta(n)$	LD $\delta(n)$
{1}	$> .028219$	$< .0049761$	10^{864}	10^{132}	$< 10^{-106}$	$< 10^{-14}$
{55,250,133}	$> .06029$	$< .0447701$	10^{208}	10^{964}	$< 10^{-24}$	$< 10^{-118}$
{136,244}	$> .024909$	$< .006398$	10^{204}	10^{420}	$< 10^{-23}$	$< 10^{-51}$
{153}	$= \frac{1}{3}$	$= \frac{1}{3}$	N/A	N/A	N/A	N/A
{160,217,352}	$> .050917$	$< .03184$	10^{160}	10^{456}	$< 10^{-18}$	$< 10^{-56}$
{370}	$> .19905$	$< .16065$	10^{276}	10^{560}	$< 10^{-32}$	$< 10^{-68}$
{371}	$> .30189$	$< .288001$	10^{836}	10^{420}	$< 10^{-102}$	$< 10^{-50}$
{407}	$> .04532$	$< .0314401$	10^{420}	10^{836}	$< 10^{-50}$	$< 10^{-103}$
{919,1459}	$> .04425$	$< .01843$	10^{916}	10^{120}	$< 10^{-112}$	$< 10^{-13}$

5.2.2. A More General Function

In order to emphasize the generality of theorem (4.1), we consider the function in base-7 with digit sequence $[0, 1, 7, 4, 17, 9, 13]$. There are the following two cycles (written in base-10) $\{1\}$, $\{20\}$, both are fixed points. The percentage of numbers arriving at 1 is large in the case, varying from .94 to .98. Figure 3 graphs the relative density of type-$\{1\}$ numbers. Table 2 shows the bounds derived. As there are only two cycles, we focus on the cycle $\{1\}$. In this case, if $n > 12$ then it satisfies bound (**B**).

Figure 2: Relative Density of Type-$\{1\}$ Integers for Digit Sequence $[0, 1, 8, 27, 64, 125, 216, 343, 512, 729]$

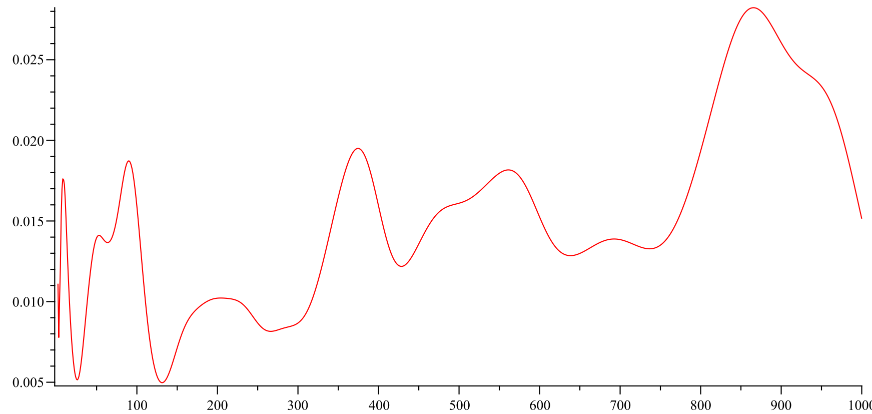
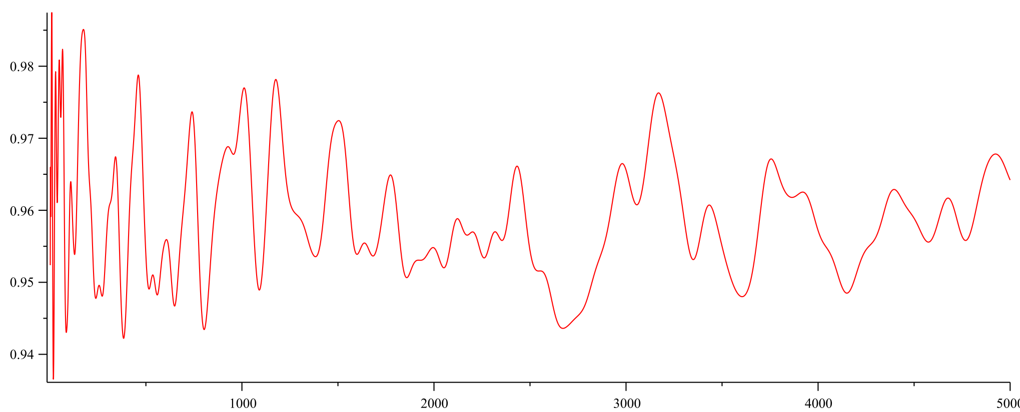


Table 2: Bounds for the Cycles Appearing for Digit Sequence $[0, 1, 7, 4, 17, 9, 13]$

Cycle	UD	LD	UD n	LD n	UD $\delta(n)$	LD $\delta(n)$
$\{1\}$	$> .9858$	$< .94222$	7^{176}	7^{384}	$< 10^{-17}$	$< 10^{-40}$

Figure 3: Relative Density of Type-$\{1\}$ Integers for Digit Sequence $[0, 1, 7, 4, 17, 9, 13]$



6. Appendix

Lemma 6.1. *Let $a > 0$, assume $f : \mathbb{R}^+ \to \mathbb{R}^+$ has continuous first and second derivatives s.t. $\forall x \in \mathbb{R}^+, f'(x) > 0$ and $f''(x) < 0$. Also we assume $\lim_{x\to\infty} f(x) = \infty$. Furthermore suppose we have $x^* \in \mathbb{R}^+$ s.t. $f(x^* + 1) \le a$. Then $\exists n \in \mathbb{N}$ s.t. $n \ge x^*$ and $0 \le a - f(n) \le f'(x^*)$.*

Proof. This follows from a first order Taylor approximation of the function f,

for completeness we have included the details. Let x^* s.t. $f(x^*+1) \leq a$ be given. Set $n = \sup\{m \in \mathbb{N} | f(m) \leq a\}$. Since f is strictly increasing and unbounded this n exists. So we have $f(n) \leq a$ and $f(n+1) > a$. Note that $n \geq x^*$ as otherwise $\lceil x^* \rceil$ would be the suprimum. By the concavity of f we have $f(n+1) - f(n) \leq f'(n) \leq f'(x^*)$. But $f(n+1) > a$, so we conclude $0 \leq a - f(n) \leq f'(x^*)$. $\square$

Lemma 6.2. *Let n be a positive integer, $\lambda = b^{\frac{n}{8}}$, and $a \in [b^{n-1}, b^n]$. Assume that μ, σ are the digit mean and variance of some generalized b-happy function H. Assume also that n satisfies the bounds in (**B**). Let $f(n) = 1 + \frac{3}{4}\mu n + \lambda\sigma\sqrt{\frac{3}{4}n}$. Then $\exists$ an integer n_2 s.t.*

- $\frac{b^{n-1}}{\mu} \leq n_2 \leq \frac{4}{3\mu} b^n$
- $4 \mid n_2$
- $0 \leq a - f(n_2) \leq 3\mu + 1$

Proof. Since we require that $4 \mid n_2$ we apply lemma 6.1 on the function $g(m) = f(4m) = 1 + 3\mu m + \lambda\sigma\sqrt{3m}$. Setting $x^* = \frac{b^{n-1}}{4\mu}$ we first check that $g(x^*+1) \leq a$. Thus we need

$$1 + 3\mu\Big(\frac{b^{n-1}}{4\mu} + 1\Big) + b^{\frac{n}{8}}\sigma\sqrt{3\Big(\frac{b^{n-1}}{4\mu} + 1\Big)} \leq a$$

Simplifying the LHS, and using the fact that $a \geq b^{n-1}$, it suffices to show

$$1 + 3\mu + b^{\frac{5n}{8}}\sigma\sqrt{\frac{3}{4b\mu} + 3b^{-n}} \leq \frac{b^{n-1}}{4}$$

To keep the results of this paper proof as general as possible, we only assumed that $\mu \geq \frac{1}{b}$ (this would correspond to the quite uninteresting b-happy function H which is $= 0$ on all digits except for 1). Also clearly $b^n \geq 3$, therefore $\frac{3}{4b\mu} + 3b^{-n} \leq \frac{3}{4} + 1 \leq 2$. And so

$$LHS \leq 1 + 3\mu + \sqrt{2}\sigma b^{\frac{5n}{8}}$$

So the statement follows if

$$4(1 + 3\mu + \sqrt{2}\sigma b^{\frac{5n}{8}}) \leq b^{n-1}$$

This is exactly the bound (**B1**). Thus by lemma 6.1 $\exists m \in \mathbb{N}$ s.t.

$$0 \leq a - g(m) \leq g'(x^*)$$

And

$$g'(x^*) = 3\mu + \frac{\sqrt{3}\sigma b^{\frac{n}{8}}}{2\sqrt{\frac{b^{n-1}}{4\mu}}}$$

$$= 3\mu + \sqrt{3\mu b}\sigma b^{\frac{-3n}{8}}$$

Again by assumption (**B2**) on n, the previous statement is

$$\leq 3\mu + 1$$

Thus setting $n_2 = 4m$ we have $4 \mid n_2, n_2 \geq \frac{b^{n-1}}{\mu}$ and $0 \leq a - f(n_2) \leq 3\mu + 1$. We note finally that $f(\frac{4}{3\mu}b^n) > a$, and since f is strictly increasing we conclude furthermore that $n_2 \leq \frac{4}{3\mu}b^n$. □

7. Acknowledgements

The author would like to thank Dr. Zeilberger, Dr. Saks, and Dr. Kopparty for their advice and support. He would also like to thank his fellow graduate students Simao Herdade and Kellen Myers for their help in editing, and fellow graduate student Patrick Devlin for generating the necessary data.

References

[1] E. El-Sedy and S. Siksek, On happy numbers. Rocky Mountain J. Math., 30 (2000), pp. 565-570

[2] H.G. Grundman, E.A. Teeple, Sequences of generalized happy numbers with small bases, J. Integer Seq. (2007) article 07.1.8

[3] Hao Pan, On consecutive happy numbers, Journal of Number Theory, Volume 128, Issue 6, June 2008, Pages 1646-1654

[4] P. Flajolet and R. Sedgewick, Analytic Combinatorics, Cambridge University Press, 2009.

[5] R.K. Guy, Unsolved Problems in Number Theory, Springer-Verlag, New York, 2nd Edition, 2004

[6] D. Moews, Personal communication, November 2011